Journal of Positive School Psychology
2022, Vol. 6, No. 3, 984–992

http://journalppw.com

Speaker Recognition for forensic application: A Review

Kavita Waghmare, Bharti Gawali

Department of CS and IT, Dr. Bababasaheb Ambedkar Marathwada University,
Aurangabad - 431 001, MS, India

Abstract

In recent years, there has been an increase in both scientific and judicial interest in forensic speaker recognition. Due to the diversity of speakers, speaker recognition is one of the most difficult challenges in biometric authentication. A forensic expert must assess evidence material during a criminal inquiry. This paper provides a brief overview of the topic of forensic speaker recognition, as well as explanations of its approaches, many feature extraction and modelling modules, applications, underlying techniques, and some performance evaluation indicators. In the overview, the limitations and application areas of existing forensic speech recognition systems are also discussed. The study concludes with a discussion of future trends and research prospects in this area.

1. Introduction

Speaker recognition is a biometric technique employed in many different contexts, with various degrees of success [1][2]. One of the most controversial usage of automatic speaker recognition is their employment in the forensics context, in which the goal is to analyze the speech data coming from wiretappings or ambient recordings retrieved during criminal investigation, with the purpose of recognizing if a given sentence had been uttered by a given person[3][4][5].

The increase in development and penetration of communication technology surely helped humankind in better, accessible and efficient communication but it is not without its ill consequences [6][7]. Information and communication technology has also helped anti-social elements in committing more organized and white collar crimes, an in turn, law enforcement agencies should be better equipped with advanced technology to counter or deal with crimes[8][9]. Speaker identification technology is one of the many tools which our law enforcement agencies could rely upon and it is also popular identification technique used for monitoring and authenticating human subjects using their speech signal[10][11].

Forensic recognition system differs from regular speaker recognition in a number of ways, including the possibility of short voice records, low voice quality, background noise, and so on [12][13][14]. The purpose of automatic speaker recognition systems is to extract, characterise, and recognise information communicating a speaker's identity from a voice sample. The task of confirming the stated identity of the speaker based on voice signal is known as forensic speaker recognition [15][16]. To determine if the unknown voice in the questioned tape belongs to the suspected speaker, a variety of procedures might be used. There is both within-speaker and between-speaker variation. As a result, forensic speaker recognition systems should produce a statistical methods that attempts to give the court an indicator of the strength of the evidence based on the estimated within-source and between-source variability[17][18].

Deep learning is a subset of machine learning that is essentially a three-layer neural network. Deep learning differs from traditional machine learning in the kind of data it uses and the learning algorithms it employs. To create predictions, machine learning algorithms use structured, labelled data that is, particular features are identified from the model's input data and grouped into tables[19][20[21].

The application of science or technology in the investigation and creation of facts or evidence in a court of law is referred to as forensic [22]. The role of forensic science is to provide knowledge to assist investigators and courts of law in answering important issues. The method of establishing if a certain individual is the source of a questioned voice recording is known as forensic speaker recognition [23]. This procedure entails comparingunknown voice recordings with one or more known voice recordings.

2. Forensic Speaker Recognition Process

Biometrics is the study of determining an individual's identification based on biological and behavioural features. Forensic automatic speaker recognition provides a data-driven biometric technology for interpreting recorded speech quantitatively as evidence. When you leave your voice as criminal evidence, a telephone recording, or an audible speech for an ear witness, forensic experts must notice the problem. Today, forensic recognition is carried out by experts, most commonly phoneticians with a linguistic and statistical background [24][25].

The method of determining if a suspected individual is the source of a questioned voice sample is known as forensic speaker recognition. The role of the forensic expert is to attest to the value of the voice evidence. The use of science and technology in the investigation and establishment of facts or evidence in a court of law is referred to as forensics [26][27]. Forensic science's role is to produce facts that can be used to prove a suspect's guilt or innocence. This process involves the comparison of recordings of an unknown voice (questioned recording) with the one or more recording of known voice (voice of the suspected speaker)
[28].

When the recognition employs any trained skill or any technologically supported process, the term technical forensic speaker recognition is often used. The approaches commonly used for technical forensic speaker recognition include the aural-perceptual, auditory-instrumental and automatic methods [29]. Forensic automatic speaker recognition is an established term used when automatic speaker recognition methods are adapted to forensic applications. In automatic speaker recognition, the deterministic or statistical models of acoustic features of the speaker’s voice and the acoustic features of questioned recordings are compared [30][31].

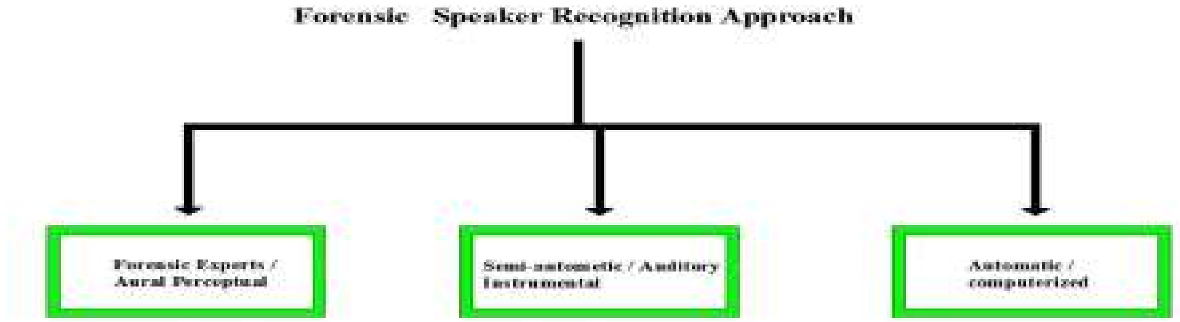

Forensic Speaker Recognition Approach
Forensic Experts / Aural Perceptual
Semi-automatic / Auditory Instrumental
Automatic / computerized

Figure1. Approach of Forensic Speaker Recognition

2.1. Aural-perceptual/Forensic Expert

Aural-perceptual approaches, which are based on human auditory perception, rely on skilled phoneticians carefully listening to recordings and estimating the level of similarity between voices based on observed variations in speech samples [32].

The ability to differentiate people by listening to their voices is one of God's gifts. Language, prosody, pitch, intensity, style, and other spectral features are used to identify a person using a range of various aspects of the human voice. There are a number of factors that an untrained listener might use to determine how to recognize a specific speaker based on these factors.

i) Identification of voice segments

ii) Detection and discrimination

iii) Linguistic content

2.2. Semi-automatic approach / Auditory-instrumental approach

Acoustic measurements of numerous variables such as the average fundamental frequency, articulation rate, formant Centre-frequencies, and other features, as well as statistical comparisons of their statistical properties, are used in auditory-instrumental approaches. The spectrographic approach of speaker 3 recognition employs a device that turns voice data into graphic representation. Spectrograms are visual representations of speech signals that transmit information about the text that the speaker has pronounced [33].

The view on similarities or dissimilarities between two specimens will be taken using this technique based on their phonetic and acoustic components such as frequencies, amplitude, plosive duration, and unvoiced signals at different positions.

2.3. Automatic Approach/ Computerized Approach

The deterministic or statistical models of acoustic aspects of the speaker's voice are compared to the acoustic features of questioned recordings using automatic methods. This strategy differs significantly from the typical approach. Feature extraction and feature modelling/classification are two stages of the process.

3. Techniques of Forensic speaker recognition

The main goal of forensic speaker recognition is to extract features from both the questioned sample and the specimen sample in order to determine whether or not they match. In forensic speaker recognition, a range of methodologies can be used. It’s analogous to a pattern recognition problem when diverse speech patterns are compared. The two following processes are involved: feature extraction and feature modelling/classification. In forensic speaker recognition, the features are first retrieved, then matched to the features of the test sample. Secondly, to characterize the speaker, a feature model is developed. The system has two phases: training and testing. In the training phase, a reference feature model is created. In the testing mode, the input signal is compared to the reference model(s) to validate or identify the speaker. The image below depicts the general framework of the forensic speaker recognition procedure [34][35][36].

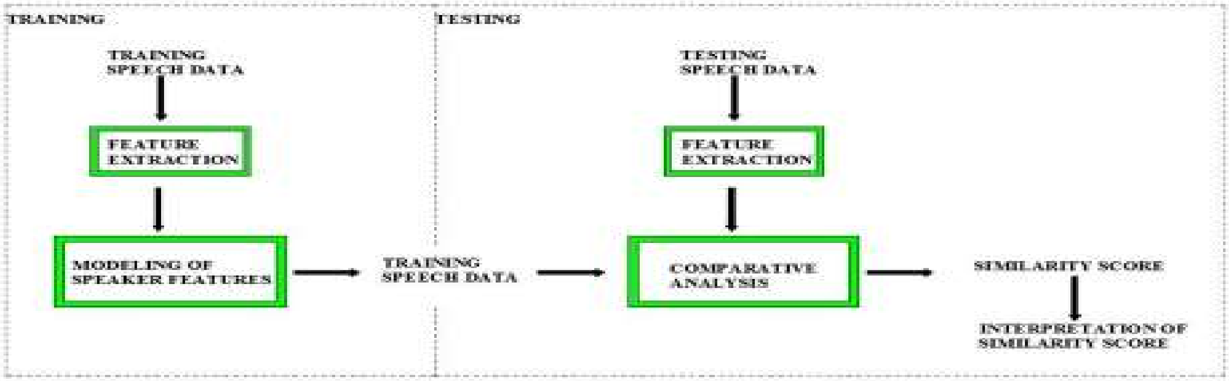


Figure 2. Forensic speaker recognition

3.1 Feature Extraction

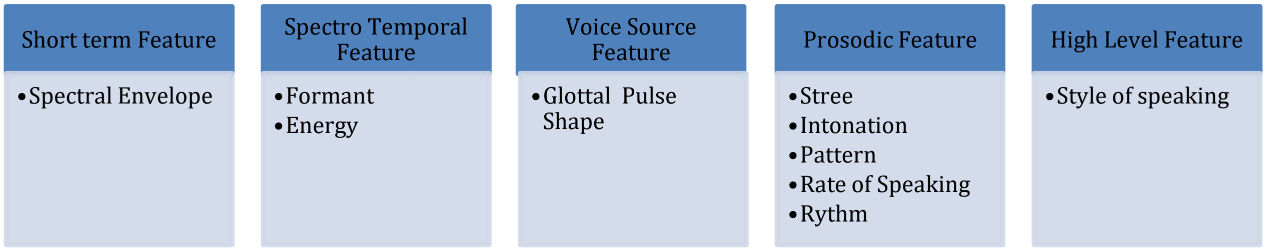
The practise of retaining relevant information in a voice signal while rejecting undesired information is known as feature extraction. The feature extraction module takes raw voice data and converts it into a feature vector with increased speaker-specific properties and statistical duplicates removed. To improve recognition performance, it is necessary to extract the optimum parametric representation of audio data. The effectiveness of this phase has an impact on the behaviour of the next phase.

The feature extraction stage's main goal is to extract speaker-specific information from voice samples for use in the identification challenge.

During the feature extraction phase, the input voice samples are converted into a sequence of multidimensional vectors, each of which corresponds to a small portion of the original speech sample [37][38].

3.1.1 Types of features

Although articulatory movements cause by the speech signal to vary constantly, the signal must be divided into brief frames. These can be classified into the following groups based on physical interpretation, as shown in figure 3.



Short term Feature	Spectro Temporal Feature	Voice Source Feature	Prosodic Feature	High Level Feature
• Spectral Envelope	• Formant • Energy	• Glottal Pulse Shape	• Stree • Intonation • Pattern • Rate of Speaking • Rythm	• Style of speaking

Figure 3 Types of speech feature

The recognition of a speaker is based on both low and high level information extracted from a voice sample. Dialect, accent, speaking style, phonetics, prosodic, and lexical information are examples of high-level information. The fundamental frequency, formant frequency, pitch, intensity, rhythm, tome spectral magnitude, and bandwidths of an individual's voice are examples of low level qualities. The following characteristics should be included in a forensic speaker recognition:

- It should be resilient against noise and distortion
- It should occur regularly and naturally in speech
- It should be easily quantified from the speech signal
- It should not be affected by the speaker's health

Forensic speaker recognition employs elements derived from a variety of speech production and perception models. Mel frequency cepstral coefficients (MFCC), Voice activity detection (VAD) and discrete wavelet transform (DWT), Vector quantization, linear predictive coding (LPC), and Linear Predictive Cepstral Coefficients (LPCC) are some of the feature extraction approaches that have been employed [39].

i) Mel-Frequency Cepstral Coefficients (MFCC)

In the literature, the most often used and popular technique for feature extraction based on the human peripheral auditory system is MFCC. It measures changes in the crucial band of the human ear using filters that are spaced linearly at low frequencies. MFCC contains two types of filters, one with linear spacing below 1000 Hz and the other with logarithmic spacing above 1000 Hz. Mel Frequency Scale includes a subjective pitch to capture crucial phonetic characteristics in speech. MFCC is made up of six crucial steps. The first step is to pre-emphasize the speech signal, which involves eliminating any undesired or noisy data. The signal is then divided into frames, which are little bits of data. The windowing of each frame is the next stage, which minimises the discontinuities at the start and end of each signal frame. The signal is then transformed from time domain to frequency domain using the Fast Fourier Transform (FFT), which reveals the spoken signal's frequencies. The signal is transmitted through the Mel-frequency wrapping block once it has been transformed to frequency domain. The Mel-objective bank's is to imitate the hearing mechanism's critical band filters. Mel-Filters focus on low frequencies while overlooking higher frequencies. Finally, the discrete cosine transform is used to log the spectrum and compress it (DCT). Mel-Frequency Cepstral Coefficients are the matrices that come from this process (MFCC). This phase creates a one-of-a-kind representation of the spectral properties of the object [40].

ii) **Voice Activity Detection (VAD)**

The term Voice Activity Detector (VAD) refers to a group of signal processing techniques that determine whether small portions of a speech signal contain voiced or unvoiced signal data. Normally, a VAD employs decision rules based on estimated signal features. VADs are used as a pre-processing block in a range of speech processing applications, including speech enhancement, speech coding, and speech and speaker recognition, when it is necessary to distinguish between voiced and unvoiced signal portions. A simple VAD works by extracting measurable features from an incoming audio signal that is separated into frames of 5-40 milliseconds in length.

The extracted features from the audio signal are then compared to a threshold limit, which is commonly calculated from the input signal's noise only periods, and a VAD judgement is made. A VAD decision (VAD = 1) is computed if the characteristic of the input frame exceeds the estimated threshold value, indicating that speech is present. If not, a VAD decision (VAD = 0) is generated, indicating that there is no speech in the input frame.

iii)Dynamic Time Warping (DTW)

Dynamic time warping is an algorithm that determines whether two sequences that differ in time or speed are similar. The varied speeds of speakers should be handled by a good ASR system. This algorithm looks for similarities between two sequences with various constraints.

IV) Vector Quantization (VQ)

Vector quantization is a type of lossy compression. The signal is first separated into vectors in vector quantization. Then, for each vector, apply quantization. VQ allows for multi-dimensional visualisation. It is possible to do so by following the procedures below:

1) A codebook is created first using a code vector.
2) Then, for each input vector in the codebook, determine the smallest Euclidean distance between them.
3) Once you've found the shortest distance, replace the vector with the codebook's index. The decision boundary is determined using the Linde BuzoGray (LBG) algorithm.

v) Linear Predictive Coding (LPC)

LPC is a digital way of encoding an analogue signal in which a specific value is anticipated by a linear function of the signal's previous values. The vocal tract, which can be thought of as a changeable diameter tube, produces human speech. The LPC model is based on a mathematical approximation of the vocal tract, which is represented by this tube of changing diameter. The speech sample s (t) is represented as a linear sum of the next sample determined by a linear combination of previous samples at any given time t. The linear nature of LPC is the most crucial feature. The value of the next sample can be determined by a linear combination of prior samples using a predictive filter [41].

vi)**Linear Predictive Cepstral Coding (LPCC)**

The biological structure of the human vocal track is shown by LPCC, which is computed via recursion from the LPC parameters to the LPC cestrum using an all pole model. The filter used in this feature extraction is usually an all-pole filter. The all-pole filter's settings. The all-pole filter's parameters are calculated using an auto-regressive approach in which the signal at each time instant can be identified using a set of preceding samples [42].

3.1.2 Feature Modelling/Classification

The speech sample is passed via the feature extraction module and the characteristic vectors are employed to create a speaker model at the time of enrolment. In speaker verification systems, a variety of modelling methodologies have been used to achieve some or all of these features. The modelling option is influenced by the type of speech to be used, the predicted performance, the ease of training and updating, as well as storage and computational variables. So will go over some of the most prevalent modelling techniques in further depth. The listener, however, is not able to understand exactly.

a) Gaussian Mixture Model

Gaussian Mixture Models (GMMs) presume that a fixed number of Gaussian distributions exist, each of which represents a cluster. As a result, the data points belonging to a single distribution tend to be grouped together in a Gaussian Mixture Model. A Gaussian mixture model (GMM) is a type of probabilistic model in which all data points are generated from a mixture of finite Gaussian distributions with unknown parameters. The parameters for Gaussian mixture models are produced from a well-trained prior model usingeither maximum a posteriori estimation or an iterative expectation-maximization approach. When it comes to modelling data, especially data from multiple groups, Gaussian mixture models are quite effective [43][44].

b) Hidden Markov Model

A hidden Markov chain is a Markov chain containing a hidden Markov. This model is analogous to a Markov model. It's a completely random process. In a Markov model, future states are decided only by the current state, not by the prior state. The states are immediately visible to the observer. It is a statistical strategy that has been effectively used in the recognition of speakers. HMM generates a statistical model of the speaker's voice production. The Viterbi technique is used to determine the likelihood of hidden states creating an unknown output sequence given the model parameters provided for the reference.

c) Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a supervised machine learning technique that can solve classification and regression problems. It is, however, mostly employed to solve categorization difficulties. Each data item is plotted as a point in n-dimensional space (where n is the number of features you have), with the value of each feature being the value of a certain coordinate in the SVM algorithm. It can be used to determine whether the data belongs to a legitimate user or an imposter. Binary SVM and multi SVM are two types of SVM. One can use binary SVM to determine whether or not a person can be identified. The attributes of two speakers are compared using binary SVM. Multi SVM, on the other hand, compares the characteristics of more than two speakers. It falls within the category of supervised classifiers.

Deep Learning

Deep learning is a subset of machine learning techniques that aims to extract high-level features from large amounts of data. It is a new area of research in many machine learning and signal processing applications. Deep Neural Network (DNN), Deep Belief Network (DBN), and Convolutional Neural Network (CNN) are several deep learning architectures that have been employed in signal processing.

Auditory analysis and spectrographic analysis are two of the most regularly employed procedures by forensic laboratories across the world. These methods are used to determine whether a criminal is guilty or innocent. The gold wave software is used by the majority of forensic science laboratories in India for pre-processing speech signals and Multi-Speech. They work using the formant frequencies f1 and f2 to determine whether the victim is guilty or innocent [45][46].

4) Database

In different languages around the world, there are a variety of options for speaker detection. The TIMIT database is used in the majority of the studies. TIMIT is a corpus of phonemically and lexically transcribed speech of different genders and dialects of American English speakers. Time has been demarcated for each transcribed element. It was created for autonomous voice and speaker recognition systems, as well as acoustic-phonetic knowledge.

The National Institute of Standards and Technology (NIST) Speaker Recognition Evaluation Database is another widely used dataset. NIST collects data to help industry, academia, and government agencies develop innovation and enhance people's lives. Other databases that have been used include Polyphone IPSC-02, which is available in French and German. PIEAS stores both telephonic and non-telephonic information. NOISEX and CSLU datasets.

5) Performance Evaluation

5.1 Accuracy Rate

There are various standards for measuring the performance of a biometric systems. Accuracy is one metric for evaluating classification models. It can also be said as the fraction of predictions. It can be given by

$$Accuracy = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

For binary classification, accuracy can also be calculated in terms of positives and negatives as follows:

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$

Where TP –True Positives, TN-True Negatives, FP- False Positives, FN-False Negatives.

5.2 Equal Error Rate (EER)

EER measures the error rate of a system when the threshold is adjusted so that the number of false acceptances is equal to the number of false rejections. The False Rejection Rate (FRR) is also known as “Type I” error. It indicates the possibility of inadvertent rejection of a person who should be able to access to the biometric system. The False Acceptance Rate (FAR) also known as “Type II” error. It shows the likelihood that someone who does not have components from a speaker and channel subspace. In a real application, the error performance can be adjusted to suit the level of security required: secure or convenient.

6) Limitations of Forensic Speaker Identification

1. Short-duration samples should be appropriately analysed.

2. Language differences are difficult to analyse.

3. Emotion variability is difficult to analyse.

4. Misspoken or misread prompted phrases.

5. Poorly recorded/noisy samples are difficult to analyse.

6. Insufficient number of comparable words.

7. Disguise in speech samples.

8. Extreme emotional states.

9. Channel mismatch during recording.

10. Different pronunciation speed of the testing and training data.

11. Variation in speech due to cough and cold.

7. Conclusion

There are still certain flaws in the speaker recognition system that can be rectified by undertaking research in sub-domains. The technology's main application is in forensic speaker recognition, where the technique’s results could be used as evidence in court. This paper provides an overview of forensic speaker recognition. Various strategies for feature extraction and modelling have been discussed. Different approaches to forensic speaker recognition have even been discussed. So there is a need for further research.

References

1. Richard D.Peacocke, Daryl H.Graf, Bell-Northern Research, “An Introduction to Speech and Speaker Recognition”, IEEE 1990.

2. J. H. L. Hansen and T. Hasan, “Speaker recognition by machines and humans: A tutorial review,” Signal Processing Magazine, IEEE, vol. 32, no. 6, pp. 74–99, 2015.

3. J. P. Campbell, W. Shen, W. M. Campbell, R. Schwartz, J. F. Bonastre, and D. Matrouf, “Forensic speaker recognition,” Institute of Electrical and Electronics Engineers, 2009.

4. A. Drygajlo, M. Jessen, S. Gfroerer, I. Wagner, J. Vermeulen, and T. Niemi, “Methodological Guidelines for Best Practice in Forensic Semiautomatic and Automatic Speaker Recognition”, Frankfurt: Verlag fur Polizeiwissenschaft, 2015.

5. Jean Francois Bonastre,Julitte Kahn, Solange RossatoMoezAjili, “ Forensics Speaker Recognition: Mirages & Reality.

6. P. Rose, Forensic Speaker Identification. London: Taylor & Francis, 2002.

7. J. P. Campbell, W. Shen, W. M. Campbell, R. Schwartz, J.-F. Bonastre, and D.Matrouf, “Forensic speaker recognition,” IEEE Signal Processing Magazine, vol. 26, no. 2, pp. 95–103, 2009.

8. F. Beritelli, “Effect of background noise on the SNR estimation of biometric parameters in forensic speaker recognition,” in Proceedings of the 2nd International Conference on Signal Processing and Communication Systems (ICSPCS ’08), IEEE, Gold Coast,

Queensland, Australia, December 2008.

9. D. A. Reynolds, “An overview of automatic speaker recognition technology,” in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, Orlando, FL, 2002, pp. 300–304.

10. SreenivasSremath Tirumala, Seyed Reza Shahamiri, "A review on Deep Learning approaches in Speaker Identification", ICSPS2016, November 21 to 24, 2016, Auckland, New Zealand.

11. S. S. Tiruala. Deep learning: Fundamentals, methods and applications. In J. Porter, editor, DEEP LEARNING USINGUNCONVENTIONALPARADIGMS, chapter 1, pages 11–. NOVA publishes, New York, 2014. 12. J. P. Campbell, W. Shen, W. M. Campbell, R. Schwartz, J.-F. Bonastre, and D.Matrouf, “Forensic speaker recognition,” IEEE Signal Processing Magazine, vol. 26, no. 2, pp. 95–103, 2009.

13. A. Drygajlo, “Automatic speaker recognition for forensic case assessment and interpretation,” Forensic Speaker Recognition: Law Enforcement and Counter-Terrorism, pp. 21–39, 2012.

14. A. Drygajlo, “Automatic speaker recognition for forensic case assessment and interpretation,” Forensic Speaker Recognition: Law Enforcement and Counter-Terrorism, pp. 21–39, 2012.

15. F. Beritelli, “Effect of background noise on the SNR estimation of biometric parameters in forensic speaker recognition,” in Proceedings of the 2nd International Conference on Signal Processing and Communication Systems (ICSPCS ’08), IEEE, Gold Coast, Queensland, Australia, December 2008.

16. T. Thiruvaran, E. Ambikairajah, and J. Epps, “FM features for automatic forensic speaker recognition,” in Proceedings of the 9th Annual Conference of the International Speech Communication Association (INTERSPEECH ’08), pp. 1497–1500, September 2008.

17. F. Denk, J. P. C. L. Da Costa, and M. A. Silveira, “Enhanced forensic multiple speaker recognition in the presence of coloured noise,” in Proceedings of the 8th International Conference on Signal Processing and Communication Systems (ICSPCS ’14), pp. 1–7, IEEE, Gold Coast, Australia, December 2014

18. Phil Rose, “Technical forensic speaker recognition: Evaluation, types and testing of evidence”, Computer Speech and Language Elsevier 2006.

19. Simon Graf, Tobias Herbig, Markus Buck, Gerhard Schmidt, “Features for voice activity detection: a comparative analysis”, Journal on Advances in Signal Processing Springer Open Access 2015. 20. LantianLi,YixiangChen,YingShi,Zhiyuan Tang, Dang Wang, “Deep Speaker Feature Learning for Text independent Speaker Verification”, INTERSPEECH 20-24 August 2017

21. Hirotaka Nakasone, Steven D. Beck, “Forensic Automatic Speaker Recognition”, A speaker Odyssey the speaker recognition workshop Crete, Greece, June 18-22 2001.

22. Roberto Togne, Daniel Pullella, “An overview of Speaker Identification: Accuracy and Robustness Issues” IEEE CIRCUITS AND SYSTEMS MAGAZINE, 27 May 2011.

23. Francesco Sigona, MirkoGrimaldi, “Tools for forensic speaker recognition” chapter 8, 2011. 24. Sergy L Koval, “Formants matching as a robust method for forensic speaker identification, SPECOM, 2006. 25. P. Rose, Forensic Speaker Identification. London: Taylor & Francis, 2002.

26. Mr.YogeshDawande, Dr.MutkaDhopeshwakar, “Analysis of different feature extraction techniques for speaker recognition system:A Review”, IJATER Vol 5,Issue 1,Jan 2005.

27. SurbhiMathur, SumitK.Choudhary, J M Vyas, “Speaker Recognition system &its forensic Implications: A Review, IJLTEMAS Vol. III, Issue IV, April 2014

28. Jose B. Trangol ,Curipe ,Abel Herrera Camacho, “Feature extraction using LPC-Residual and Mel-Frequency Cepstral Coefficients in Forensic Speaker Recognition, IJCEE, Vol.5,No.1, Feb 2013. 29. AmnaIrum,Ahmad Salman, “Speaker Verification using Deep Neural Networks: A review”, IJMLC,Vol. 9, No.1, Feb 2019.

30. SupapornBunrit, ThuttapholInkian, NittayaKerdprsop,KittasakKerdprasop, “Text-Independent speaker Identification using Deep learning Model of Convolution Neural Network, IJMLC, Vol.9, No.2,April 2019

31. Ahmed Kamil, Hasan Al-Ali,DavidDean,BouchrSenadji, Vinod Chandran, Ganesh. R.Naik, “Enhanced Forensic Speaker Verification using a combination of DWT & MFCC feature warping in the presence of Noise & Reverberation Condition” 22 Aug 2017.

32. Ewald Enzinger,Geoffray Stewart Morrison, “Empirical test of the performance of an acoustic –phonetic approach to forensic voice comparison under conditions similar to those of a real case, Elsevier 2017. 33. Harriet M.J.Smith,ThomS.Bagulay, Jeremy Robson, Andrew K.Dumn,PaulaC.Stacey, “Forensic voice discrimination by lay listeners: The effect of speech type and background noise on performance”, John Wiley & sons Ltd 2018.

34. Andrey Barinov, Sergey Koval, Pavel Ignatov, Mikhail,Stollov,“Channel Compensation for Forensic Speaker Identification using inverse processing, AES 39th International Conference.

35. Nguyen Nang An, Nguyen Quang Thanh, Yanbing Liu, “Deep CNNs with Self-attention for Speaker Identification”, IEEE Access 15 July, 2019.

36. BabitaBhall, Singh CP, Rakesh Dhar, Rajesh Soni, “Auditory & Acoustic Features from Clue-Words sets for Forensic Speaker Identification & its correlation

with Probability Scales" JFP 2016. 37. Elizabeth Shriberg, Andreas Stolcke, "The case of Automatic Higher- Level Features in Forensic Speaker Recognition, Proceedings of the 9th International Conference of the ISCA 2008.

38. CuilingZhang,Joost van de Weijer, Jingxu Cui, "Intra and inter-speaker variations of formant pattern for lateral syllables in Standard Chinese", Elsevier Forensic Science International 2006. 39. Eugenia San Segundo, Athanasios Tsanas, Pedro Gomez-Vilde, "Euclidean Distances as measures of speaker similarity including identical twin pairs: A forensic investigation using source and filter voice characteristic", Elsevier2017.

40. Francesco Beritelli,AndreaSpadaccini, "Performance Evaluation of Automatic Speaker Recognition Technique for Forensics Applications", Intech 2012.

41. Satyanand Singh, "Forensic and Automatic Speaker Recognition System", IJECE Vol.8, No.5, October2018. 42. Volker Dellwo, AdrainLeemann, Marie-Jose Kolly, "The recognition of read and spontaneous speech in local vernacular: The case of Zurich German", Elsevier 8 OCT 2014.

43. Phil Rose, "Forensic Voice Comparison with Japanese Vowel Acoustics-A Likelihood Ratio-Based Approach using segmental cepstra", ICPhS XVII Hong Kong 17-21 August 2011.

44. David Sztaho, GyorgySzaszak, Andras Beke, "Deep learning methods in speaker recognition: A review", Audio and Speech Processing 14 Nov 2019.

45. Yun Lei, Luka´s Burget, and Nicolas Scheffer, "A noise robust i-vector extractor using vector taylorseries for speaker recognition," in Proc. of ICASSP, 2013, pp. 6788–6791.

46. A. Eriksson, "Tutorial on forensic speech science. Part I: Forensic phonetics", Proceedings of the 9th European Conference on Speech Communication and Technology, Lisbon, Sept. 4-8, 2005.